Probabilistic Analysis of Onion Routing in a Black-box Model

[Extended Abstract] *

Joan Feigenbaum[†]
Yale University

Aaron Johnson[‡]
Yale University

Paul Syverson[§]
Naval Research Laboratory

ABSTRACT

We perform a probabilistic analysis of onion routing. The analysis is presented in a black-box model of anonymous communication that abstracts the essential properties of onion routing in the presence of an active adversary that controls a portion of the network and knows all *a priori* distributions on user choices of destination. Our results quantify how much the adversary can gain in identifying users by exploiting knowledge of their probabilistic behavior. In particular, we show that a user u's anonymity is worst either when the other users always choose the destination u is least likely to visit or when the other users always choose the destination u chooses. This worst-case anonymity with an adversary that controls a fraction b of the routers is comparable to the best-case anonymity against an adversary that controls a fraction $\sqrt{b}$.

Categories and Subject Descriptors

C.2.0 [**Computer-Communication Networks**]: General—*Security and protection*; C.2.4 [**Computer-Communication Networks**]: Distributed Systems—*Distributed applications*

General Terms

Algorithms, Security, Theory

Keywords

Anonymity, onion routing, Tor

1. INTRODUCTION

Every day, thousands of people use the onion-routing network Tor [6] to anonymize their Internet communication. However, the effectiveness of this service, and of onion routing in general, is not well understood. The approach we take to this problem is to model onion routing formally all the way from the protocol details to the behavior of the users. We then analyze the resulting system and quantify the anonymity it provides. Key features of our model include *i*) a black-box abstraction that hides the underlying operation of the protocol and *ii*) probabilistic user behavior and protocol operation.

Systems for communication anonymity generally have at most one of two desirable properties: provable security and practicality. Systems that one can prove secure require assumptions that make them impractical for most communication applications. Practical systems are ultimately the ones we must care about, because they are the ones that will actually be used. However, their security properties have not been rigorously analyzed or even fully stated. This is no surprise, because practical anonymity systems have been deployed and available to study for perhaps a decade, while practical systems for communications confidentiality and/or authenticity have been in use almost as long as there have been electronic communications. It often takes a while for theory and practice to catch up to each other.

Of the many anonymous-communication design proposals, onion routing [9] has had notable success in practice. Several implementations have been made [9, 27, 6], and there was a similar commercial system, Freedom [1]. As of January 2007, the most recent iteration of the basic design, Tor [6], consists of over 800 routers, collectively processing an average of 100MB/s, with an estimated total user population of 200,000 [17]. Because of this popularity, we believe it is important to improve our understanding of the protocol.

Onion routing is a practical anonymity-network scheme with relatively low overhead and latency. It provides two-way, connection-based communication and does not require that the destination participate in the anonymity-network protocol. These features make it useful for anonymizing much of the communication that takes place over the Internet today, such as web browsing, chatting, and remote login. Thus, formal analysis and provable anonymity results for onion routing are significant

In a recent paper [7], we took a first step toward bridging the gap between provability and practicality by presenting a formal I/O-automata model of onion routing and proving anonymity guarantees. In the present paper, we ab-

[*] A full version of this paper has been submitted for journal publication and is available in preprint form from the authors.

[†] Supported in part by NSF grants 0331548 and 0534052, ARO grant W911NF-06-1-0316, and US-Israeli BSF grant 2002065. Email: joan.feigenbaum@yale.edu

[‡] Supported by NSF grant 0428422 and ARO grant W911NF-05-1-0417. Email: aaron.johnson@yale.edu

[§] Supported by ONR. Email: syverson@itd.nrl.navy.mil

WPES'07 October 29, 2007, Alexandria, Virginia, USA
Copyright 2007 ACM 978-1-59593-883-1/07/0010 ...$5.00.

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **2007**	2. REPORT TYPE	3. DATES COVERED **00-00-2007 to 00-00-2007**

4. TITLE AND SUBTITLE **Probabilistic Analysis of Onion Routing in a Black-box Model**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,Code 7215, Remote Sensing Division,4555 Overlook Avenue, SW,Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

WPES’07: Proceedings of the 2007 ACM Workshop on Privacy in Electronic Society, ACM Press, October 2007, pp. 1-10

14. ABSTRACT

We perform a probabilistic analysis of onion routing. The analysis is presented in a black-box model of anonymous communication that abstracts the essential properties of onion routing in the presence of an active adversary that controls a portion of the network and knows all a priori distributions on user choices of destination. Our results quantify how much the adversary can gain in identifying users by exploiting knowledge of their probabilistic behavior. In particular, we show that a user u’s anonymity is worst either when the other users always choose the destination u is least likely to visit or when the other users always choose the destination u chooses. This worst-case anonymity with an adversary that controls a fraction b of the routers is comparable to the bestcase anonymity against an adversary that controls a fraction pb.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**10**	

stract away from that model and treat the network simply as a black box to which users connect and through which they communicate with destinations. The abstraction captures the relevant properties of a protocol execution that the adversary can infer from his observations - namely, the observed users, the observed destinations, and the possible connections between the two. In this way, we abstract away from much of the design specific to onion routing so that our results apply both to onion routing and to other low-latency anonymous-communication designs. The executions described in [7] disappear from the analysis done herein.

Our previous analysis in the I/O-automata model was possibilistic, a notion of anonymity that is simply not sensitive enough. It makes no distinction between communication that is equally likely to be from any one of a hundred senders and communication that came from one sender with probability .99 and from each of the other 99 senders with probability .000101. An adversary in the real world is likely to have information about which scenarios are more realistic than others. In particular, users' communication patterns are not totally random. When the adversary can determine with high probability, *e.g.*, the sender of a message, that sender is not anonymous in a meaningful way.

Using this intuition, we add a probability measure to the I/O-automata model of [7]. For any set of actual circuit sources and destinations, there is a larger set that is consistent with the observations made by an adversary. The adversary can then infer conditional probabilities on this larger set using the measure. This gives the adversary probabilistic information about the facts we want the network to hide, such as the initiator of a communication.

The probability measure that we use models heterogeneous user behavior. In the onion-routing protocol specified in [7], each user chooses a circuit to a destination. We make this choice probabilistic and have each user choose a destination according to some probability distribution, allowing this distribution to be different for different users. We assume that the users choose their circuits by selecting the routers on it independently and at random.

After observing the protocol, the adversary can in principle infer some distribution on circuit source and destination. He may not actually know the underlying probability measure, however. In particular, it doesn't seem likely that the adversary would know how every user selects destinations. In our analysis, we take a worst-case view and assume that the adversary knows the distribution exactly. Also, over time he might learn a good approximation of user behavior via the long-term intersection attack [4]. In this case, it may seem as though anonymity has been essentially lost anyway. However, even when the adversary knows how a user generally behaves, the anonymity network may make it hard for him to determine who is responsible for any specific action, and the anonymity of a specific action is what we are interested in.

We analyze *relationship anonymity* [19, 25] in our onion routing model. Relationship anonymity is obtained when the adversary cannot identify the destination of a user. In terms of the conventional subject/action specification for anonymity [19], we can take the action to be communication from a given user u and the subject to be the destination. Suggested probabilistic metrics for anonymity applied to this case include probability assigned to the correct destination [21], the entropy of the destination distribution [5, 23], and maximum probability within the destination distribution [29]. We will use the probability assigned to the correct destination as our metric. In part, this is because it is the simplest metric. Also, any statements about entropy and maximum probability metrics only make loose guarantees about the probability assigned to the actual subject, a quantity that clearly seems important to the individual users.

We look at the value of this anonymity metric for a choice of destination by a user. Fixing a destination by just one user, say u, does not determine what the adversary sees, however. The adversary's observations are also affected by the destinations chosen by the other users and the circuits chosen by everybody. Because those variables are chosen probabilistically under the measure we added, the anonymity metric will have its own distribution. Several statistics about this distribution might be interesting; in this paper, we look at its expectation.

The distribution of the anonymity metric for a given user and destination depends on the other users' destination distributions. If their distributions are very different, the adversary may have an easy time separating out the actions of the user. If they are similar, the user may more effectively hide in the crowd.

We begin by providing a kind of worst-case guarantee to a user with a given destination distribution by finding the maximum expectation over the possible destination distributions of the other users. Our results show that the worst case is when every other user either always visits the destinations the user is otherwise least likely to visit or always visits his actual destination. Which one is worse depends on how likely he was to visit his destination in the first place. If he is unlikely to visit it, it is worse when everybody else always visits his otherwise least-likely destination, because the adversary can generally infer that he is not responsible for communication to that destination. When he is likely to visit it, the adversary considers him likely to be the culprit whenever the destination is observed, and so observing that destination often causes the adversary to suspect the truth. We give an approximation to the user's anonymity in these worst cases for large user populations that shows that on average it decreases by about the fraction of the network the adversary controls.

We then consider anonymity in a more typical set of user distributions. In the model suggested by Shmatikov and Wang [25], each user selects a destination from a common Zipfian distribution. Because the users are identical, every user hides well among the others. As the user population grows, the anonymity loss in this case tends to the square of the fraction of the network that is compromised.

1.1 Related Work

Ours is not the first formalization of anonymous communication. Earlier formalizations used CSP [22], graph theory and possible worlds [11], and epistemic logic [28, 10]. These earlier works focused primarily on formalizing the high-level concept of anonymity in communication. For this reason, they applied their formalisms to toy examples or systems that are of limited practical application and can only provide very strong forms of anonymity, e.g., dining-cryptographers networks. Also, with the exception of [10], they have at most a limited ability to represent probability and probabilistic reasoning. We have focused in [7] on formalizing a widely deployed and used, practical, low-latency system.

Halpern and O'Neill [10] give a general formulation of anonymity in systems that applies to our model. They describe a "runs-and-systems" framework that provides semantics for logical statements about systems. They then give several logical definitions for varieties of anonymity. It is straightforward to apply this framework to the network model and protocol that we give in [7]. Our possibilistic definitions of sender anonymity, receiver anonymity, and relationship anonymity then correspond to the notion of "minimal anonymity" as defined in their paper. The other notions of anonymity they give are generally too strong and are not achieved in our model of onion routing.

Even earlier formalizations of substantial anonymous communication systems [2, 16] have not been directly based on the design of deployed systems and have focused on provability without specific regard for applicability to an implemented or implementable design. Also, results in both of these papers are for message-based systems, rather than systems that create a cryptographic circuit prior to communication. Thus, while illuminating, they are not likely to be applicable to low-latency communications, and, despite the title of [2], are not analyses of onion routing.

In this paper, we add probabilistic analysis to the framework of [7]. Other works have presented probabilistic analysis of anonymous communication [21, 24, 30, 3, 4, 15, 13] and even of onion routing [27]. The work of Shmatikov and Wang [25] is particularly similar to ours. It calculates relationship anonymity in mix networks and incorporates user distributions for selecting destinations. However, with the exception of [24], these have not been formal analyses. Also, whether for high-latency systems such as mix networks, or low-latency systems, such as Crowds and onion routing, many of the attacks in these papers are some form of intersection attack. In an intersection attack, one watches repeated communication events for patterns of senders and receivers over time. Unless all senders are on and sending all the time (in a way not selectively blockable by an adversary) and/or all receivers receiving all the time, if different senders have different receiving partners, there will be patterns that arise and eventually differentiate the communication partners. It has long been recognized that no system design is secure against a longterm intersection attack. Several of these papers set out frameworks for making that more precise. In particular, [3], [4], and [15] constitute a progression towards quantifying how long it takes (in practice) to reveal traffic patterns in realistic settings.

We are not concerned herein with intersection attacks. We are effectively assuming that the intersection attack is done. The adversary already has a correct distribution of a user's communication partners. We are investigating the anonymity of a communication in which a user communicates with one of those partners in the distribution. This follows the anonymity analyses performed in much of the literature [13, 16, 21, 27], which focus on finding the source and destination of an individual communication. Our analysis differs in that we take into account the probabilistic nature of the users' behavior.

We expect this to have potential practical applications. For example, designs for shared security-alert repositories to facilitate both forensic analysis for improved security design and quicker responses to widescale attacks have been proposed [14]. A participant in a shared security-alert repository might expect to be known to communicate with it on a regular basis. Assuming reports of intrusions, etc. are adequately sanitized, the concern of the participant should be to hide when it is that updates from that participant arrive at the repository, *i.e.*, which updates are likely to be from that participant as opposed to others.

2. TECHNICAL PRELIMINARIES

2.1 Model

We describe our analysis of onion routing in terms of a black-box model of anonymous communication. We are using a black-box model for two reasons: First, it abstracts away the nonessential details, and second, its generality immediately suggests ways to perform similar analyses of other anonymity networks. It models a round of anonymous communication as a set of inputs owned by users and a set of outputs owned by destinations. The adversary observes the source of some of the inputs and the destination of some of the outputs. This captures the basic capabilities of an adversary in an onion-routing network that controls some of the routers. In this situation, the adversary can determine the source of messages when he controls the first router on the source's circuit and the destination of messages when he controls the last router. In order for the adversary always to be able to recognize when it controls the onion router adjacent to the circuit source, we assume that the initiating client is not located at an onion-routing network node. This is the case for the vast majority of circuits in Tor and in all significant deployments of onion routing and similar systems to date. We discuss this assumption further in section 5. The black box system can also model a mix network under attack by a global, passive adversary. Such a model was used by Kesdogan et al. [12] in their analysis of an intersection attack.

We add two assumptions to specialize this model to onion routing. First, we assume that every user owns exactly one input and is responsible for exactly one output in a round. Certainly users can communicate with multiple destinations simultaneously in actual onion-routing systems. However, it seems likely that in practice most users have at most some small constant number of active connections at any time, and the smaller this constant is the fewer possibilities there are that are consistent with the adversary's observations. Therefore, this assumption is a conservative one that gives the adversary as much power to break anonymity as the limited number of user circuits can provide. Second, we assume the adversary can link together an input and output from the same user when he observes them both. This is another conservative assumption that is motivated by the existence of timing attacks that an active adversary can use to link traffic that it sees at various points along its path through the network.

Let U be the set of users with $|U| = n$. Let Δ be the set of destinations. A round of communication C in a black-box system is defined by a selection of a destination by each user, $C_D : U \to \Delta$, a set of users whose inputs are observed, $C_I : U \to \{0, 1\}$, and a set of users whose outputs are observed, $C_O : U \to \{0, 1\}$. The round C will also be referred to as a *configuration.* A user's input, output, and destination will be called its *circuit.*

We include the probabilistic behavior of users by adding a probability measure over configurations. Let each user u

select a destination d from a distribution p^u over Δ, where we denote the probability that u chooses d as p^u_d. Every input and output is independently observed with probability b. This reflects the probability that the first or last router of a user's circuit is compromised when the user selects the circuit's routers independently and at random, and the adversary controls a fraction b of the routers. The probability of a configuration C is the joint probability of its events:

$$Pr[C] = \prod_{u \in U} \left(p^u_{C_D(u)}\right) \left(b^{C_I(u)}(1-b)^{1-C_I(u)}\right) \cdot \left(b^{C_O(u)}(1-b)^{1-C_O(u)}\right) \tag{1}$$

For any configuration, there is a larger set of configurations that are consistent with the inputs and outputs that the adversary sees. We will call two configurations *indistinguishable* if the sets of inputs, outputs, and links between them that the adversary observes are the same.

Definition 1. Configurations C and $\overline{C}$ are *indistinguishable* if there exists a permutation $\pi : U \to U$ such that for all $u \in U$:

1. $C_I(u) = 1 \wedge C_O(u) = 1 \Rightarrow$ $\overline{C}_I(u) = 1 \wedge \overline{C}_O(u) = 1 \wedge C_D(u) = \overline{C}_D(u)$
2. $C_I(u) = 1 \wedge C_O(u) = 0 \Rightarrow$ $\overline{C}_I(u) = 1 \wedge \overline{C}_O(u) = 0$
3. $C_I(u) = 0 \wedge C_O(u) = 1 \Rightarrow$ $\overline{C}_I(\pi(u)) = 0 \wedge \overline{C}_O(\pi(u)) = 1 \wedge C_D(u) = \overline{C}_D(\pi(u))$
4. $C_I(u) = 0 \wedge C_O(u) = 0 \Rightarrow$ $\overline{C}_I(\pi(u)) = 0 \wedge \overline{C}_O(\pi(u)) = 0$

Thus, two configurations are indistinguishable if they have the same pattern of observed inputs, outputs, and destinations, while allowing the identities of users with unobserved inputs to be permuted. The adversary relation is an equivalence relation, and, in particular, is symmetric, because, if C and $\overline{C}$ are indistinguishable under π, then $\overline{C}$ and C are indistinguishable under π^{-1}. Therefore we use the notation $C \approx \overline{C}$ to indicate that configurations C and $\overline{C}$ are indistinguishable.

2.2 Probabilistic Anonymity

A user performs an action anonymously in a possibilistic sense if there is an indistinguishable configuration in which the user does not perform the action. For example, under this definition a user with observed output but unobserved input sends that output anonymously if there exists another user with unobserved input. The probability measure we have added to configurations allows us to incorporate the degree of certainty that the adversary has about the subject of an action. After making observations in the actual configuration, the adversary can infer a conditional probability distribution on configurations. There are several candidates in the literature for assessing an anonymity metric from this distribution. The probabilistic anonymity metric that we use is the posterior probability of the correct subject. The lower this is, the more anonymous we consider the user.

We note that the calculation of posterior probabilities in the black-box model carries down to the I/O-automata model of [7]. Under a notion of indistinguishability of executions, an assumption that all executions of a configuration are equally likely, and an assumption that the adversary does not block any circuits, the distribution of the posterior probability of an otherwise active adversary for any user action that depends only on a configuration (*e.g.* that of relationship anonymity) is identical to the one in the associated black-box model. Thus, the black box is a valid abstraction of our formal model of onion routing, and the results we derive for it hold in that more detailed model.

2.3 Relationship Anonymity

We analyze the relationship anonymity of users and destinations in our model. We measure the relationship anonymity of user u and destination d by the posterior probability that u chooses d as his destination. The lower this is, the more anonymous we consider their relationship.

The relationship anonymity of u and d varies with the destination choices of the other users and the observations of the adversary. If, for example, u's output is observed, and the inputs of all other users are observed, then the adversary knows u's destination with probability 1. Because we want to examine the relationship anonymity of u conditioned only on his destination, we end up with a distribution on the anonymity metric. We look at the expectation of this distribution. Moreover, because this distribution depends on the destination distributions of all of the users, we continue by finding the worst-case expectation for a given user and destination and then examine the expectation in a more likely situation.

3. EXPECTED ANONYMITY

Because of space limitations, full proofs of some of the results in this sections are omitted. They are included in the journal submission, which is available in preprint form from the authors.

Let the set $\mathcal{C}$ of all configurations be the sample space and X be a random configuration. X is then distributed according to Equation 1. Let Y be the posterior probability of the event that u chooses d as a destination, that is, $Y(C) = Pr[X_D(u) = d | X \approx C]$. Y is our metric for the relationship anonymity of u and d.

3.1 Calculation and Bounds

Let $\mathbb{N}^\Delta$ represent the set of multisets over Δ. Let $\Pi(A, B)$ be the set of all injective maps $A \to B$. Let $\rho(\Delta^0)$ be the number of permutations of $\Delta^0 \in \mathbb{N}^\Delta$ that only permute elements of the same type:

$$\rho(\Delta^0) = \prod_{\delta \in \Delta} |\{\delta \in \Delta^0\}|!$$

The following theorem gives an exact expression for the conditional expectation of Y in terms of the underlying parameters U, Δ, p, and b:

Theorem 1.

$$E[Y|X_D(u) = d] = b(1-b)p_d^u + b^2 + \sum_{S\subseteq U: u\in S} \sum_{\Delta^0\in\mathbb{N}^\Delta:|\Delta^0|\leq S} b^{n-|S|+|\Delta^0|}(1-b)^{2|S|-|\Delta^0|} \cdot \left(\sum_{T\subseteq S-u:|T|=|\Delta^0|-1} \sum_{\pi\in\Pi(T+u,\Delta^0):\pi(u)=d} p_d^u \prod_{v\in T} p_{\pi(v)}^v + \sum_{T\subseteq S-u:|T|=|\Delta^0|} \sum_{\pi\in\Pi(T,\Delta^0)} p_d^u \prod_{v\in T} p_{\pi(v)}^v \right)^2 \cdot [\rho(\Delta^0)]^{-1}(p_d^u)^{-1} \left(\sum_{T\subseteq S:|T|=|\Delta^0|} \sum_{\pi\in\Pi(T,\Delta^0)} \prod_{v\in T} p_{\pi(v)}^v \right)^{-1} \quad (2)$$

Proof. At a high level, the conditional expectation of Y can be expressed as:

$$\mathrm{E}[Y|X_D(u) = d] = \sum_{C\in\mathcal{C}} \Pr[X = C|X_D(u) = d]Y(C)$$

We calculate Y for a configuration C by finding the relative weight of indistinguishable configurations in which u selects d. The adversary observes some subset of the circuits. If we match the users to circuits in some way that sends users with observed inputs to their own circuits, the result is an indistinguishable configuration. Similarly, we can match circuits to destinations in any way that sends circuits on which the output has been observed to their actual destination in C.

The value of $Y(C)$ is especially simple if u's input has been observed. If the output has not also been observed, then $Y(C) = p_d^u$. If the output has also been observed, then $Y(C) = 1$.

For the case in which u's input has not been observed, we have to take into account the destinations of and observations on the other users. Let $S \subseteq U$ be the set of users s such that $C_I(s) = 0$. Note that $u \in S$. Let Δ^0 be the multiset of the destinations of circuits in C on which the input has not been observed, but the output has.

Let $f_0(S, \Delta^0)$ be the probability that in a random configuration the set of unobserved inputs is S and the set of observed destinations with no corresponding observed input is Δ^0:

$$f_0(S,\Delta^0) = b^{n-|S|+|\Delta^0|}(1-b)^{2|S|-\Delta^0}[\rho(\Delta)]^{-1} \cdot \sum_{T\subseteq S:|T|=|\Delta^0|} \sum_{\pi\in\Pi(T,\Delta^0)} \prod_{v\in T} p_{\pi(v)}^v$$

Let $f_1(S, \Delta^0)$ be the probability that in a random configuration the set of unobserved inputs is S, the set of observed destinations with no corresponding observed input is Δ^0, the output of u is observed, and the destination of u is d:

$$f_1(S,\Delta^0) = b^{n-|S|+|\Delta^0|}(1-b)^{2|S|-\Delta^0}[\rho(\Delta)]^{-1}p_d^u \cdot \sum_{T\subseteq S-u:|T|=|\Delta^0|-1} \sum_{\pi\in\Pi(T+u,\Delta^0):\pi(u)=d} \prod_{v\in T} p_{\pi(v)}^v$$

Let $f_2(S, \Delta^0)$ be the probability that in a random configuration the set of unobserved inputs is S, the set of observed destinations with no corresponding observed input is Δ^0, the output of u is unobserved, and the destination of u is d:

$$f_2(S,\Delta^0) = b^{n-|S|+|\Delta^0|}(1-b)^{2|S|-\Delta^0}[\rho(\Delta)]^{-1}p_d^u \cdot \sum_{T\subseteq S-u:|T|=|\Delta^0|} \sum_{\pi\in\Pi(T,\Delta^0)} \prod_{v\in T} p_{\pi(v)}^v$$

Now we can express the posterior probability $Y(C)$ as:

$$Y(C) = \frac{f_1(S,\Delta^0) + f_2(S,\Delta^0)}{f_0(S,\Delta^0)} \quad (3)$$

The expectation of Y is a sum of the above posterior probabilities weighted by their probability. The probability that the input of u has been observed but the output hasn't is $b(1-b)$. The probability that both the input and output of u have been observed is b^2. These cases are represented by the first two terms in Equation 2.

When the input of u has not been observed, we have an expression of the posterior in terms of sets S and Δ^0. The numerator $(f_1 + f_2)$ of Equation 3 itself actually sums the weight of every configuration that is consistent with S, Δ^0, and the fact that the destination of u is d. However, we must divide by p_d^u, because we condition on the event $\{X_D(u) = d\}$.

These observations give us the final summation in Equation 2. □

The expression for the conditional expectation of Y in Equation 2 is complicated and unenlightening. It would be nice if we could find a simple approximation. The probabilistic analysis in [27] proposes just such a simplification by reducing it to only two cases: *i)* the adversary observes the user's input and output and therefore identifies his destination and *ii)* the adversary doesn't observe these and cannot improve his *a priori* knowledge. The corresponding simplified expression for the expection is:

$$E[Y|X_D(u) = d] \approx b^2 + (1-b^2)p_d^u \quad (4)$$

This is a reasonable approximation if the final summation in Equation 2 is about $(1-b)p_d^u$. This summation counts the case in which u's input is not observed, and to achieve a good approximation the adversary must experience no significant advantage or disadvantage from comparing the users with unobserved inputs (S) with the discovered destinations (Δ^0).

The quantity $(1-b)p_d^u$ does provide a lower bound on the final summation. It may seem obvious that considering the destinations in Δ^0 can only improve the accuracy of adversary's prior guess about u's destination. However, in some situations the posterior probability for the correct destination may actually be smaller than the prior probability. This may happen, for example, when some user v, $v \neq u$, communicates with a destination e, $e \neq d$, and only u is *a priori* likely to communicate with e. If the adversary observes the communication to e, it may infer that it is likely that u was responsible and therefore didn't choose d.

It is true, however, that in expectation this probability can only increase. Therefore Equation 4 provides a lower bound on the expected anonymity.

Theorem 2. *$E[Y|X_D(u) = d] \geq b^2 + (1-b^2)p_d^u$*

Proof Sketch. As described in the proof of Theorem 1,

$$E[Y|X_D(u) = d] = b^2 + b(1-b)p_d^u + (1-b)E[Y|X_D(u) = d \wedge X_I(u) = 0]$$

An application of the Cauchy-Schwartz inequality shows that $E[Y|X_D(u) = d \wedge X_I(u) = 0] \geq p^u_d$. □

To examine the accuracy of our approximation, we look at how large the final summation in Equation 2 can get as the users' destination distributions vary. Because this is the only term that varies with the other user distributions, this will also provide a worst-case guarantee on expected anonymity. Our results will show that the worst case can occur when the users other than u act as differently from u as possible by always visiting the destination u is otherwise least likely to visit. Less obviously, we show that the maximum can also occur when the users other than u always visit d. This happens because it makes the adversary observe destination d often, causing him to suspect that u chose d. Our results also show that the worst-case expectation is about $b + (1-b)p^u_d$, which is significantly worse than the simple approximation above.

As the first step in finding the maximum of Equation 2 over $(p^v)_{v \neq u}$, we observe that it is obtained when every user $v \neq u$ chooses only one destination d_v, *i.e.* $p^v_{d_v} = 1$ for some $d_v \in \Delta$.

LEMMA 1. *A maximum of* $E[Y|X_D(u) = d]$ *over* $(p^v)_{v \neq u}$ *must occur when, for all* $v \neq u$, *there exists some* $d_v \in \Delta$ *such that* $p^v_{d_v} = 1$.

PROOF. Take some user $v \neq u$ and two destinations $e, f \in \Delta$. Assign arbitrary probabilities in p^v to all destinations except for f, and let $\zeta = 1 - \sum_{\delta \neq e,f} p^v_\delta$. Then $p^v_f = \zeta - p^v_e$. Consider $E[Y|X_D(u) = d]$ as a function of p^v_e. The terms t_i of $E[Y|X_D(u) = d]$ that correspond to any fixed S and Δ^0 are of the following general form:

$$t_i = \frac{(\alpha_i p^v_e + \beta_i(\zeta - p^v_e) + \gamma_i)^2}{\delta_i p^v_e + \epsilon_i(\zeta - p^v_e) + \eta_i}$$

This is a convex function of p^v_e:

$$t''_i = \frac{2(\gamma_i(\delta_i - \epsilon_i) + \beta_i(\delta_i\zeta + \eta_i) - \alpha_i(\epsilon_i\zeta + \eta_i))^2}{(\epsilon_i(\zeta - p^v_e) + \delta_i p^v_e + \eta_i)^3} \geq 0$$

The leading two terms of $E[Y|X_D(u) = d]$ are constant in p^v, and the sum of convex functions is a convex function, so $E[Y|X_D(u) = d]$ is convex in p^v_e. Therefore, a maximum of $E[Y|X_D(u) = d]$ must occur when $p^v_e \in \{0, 1\}$. □

The following lemma shows that we can further restrict ourselves to distribution vectors in which, for every user except u, the user either always chooses d or always chooses the destination that u is otherwise least likely to visit.

LEMMA 2. *Order the destinations* $d = d_1, \ldots, d_{|\Delta|}$ *such that* $p^u_{d_i} \geq p^u_{d_{i+1}}$ *for* $i > 1$. *Then a maximum of* $E[Y|X_D(u) = d]$ *must occur when, for all users* v, *either* $p^v_{d_1} = 1$ *or* $p^v_{d_{|\Delta|}} = 1$.

PROOF SKETCH. Assume, following Lemma 1, that $(p^v)_{v \neq u}$ is an extreme point of the set of possible distribution vectors. Let $S \ni u$ be the set of users with unobserved inputs and $T \subseteq S$ be those that further have observed outputs. Let $s_{d_k} = |\{s \in S : p^u_{d_k} = 1\}|$. For $1 < i < j$, let $m = s_{d_i} + s_{d_j}$. Holding m constant and trading off s_{d_i} and s_{d_j}, the expression for the expectation of Y conditioned on S and T is maximized when $s_{d_i} = 0$. Therefore, a weighted sum over all such sets S and T is maximized when $|\{v \in U - u : p^v_{d_i} = 1\}| = 0$. $E[Y|X_D(u) = d \wedge X_O(u) = 0]$ is equal to such a sum. Equation 2 shows that, when $X_O(u) = 1$, the conditional expectation doesn't vary with the destination distributions p^v. □

Therefore, in looking for a maximum we can assume that every user except u either always visits d or always visits $d_{|\Delta|}$. We can use the same idea with d and $d_{|\Delta|}$ as was used with d_i and d_j and consider $E[Y|X_D(u) = d]$ as we trade off the number of users visiting them. Doing this shows that a maximum is obtained either when all users but u always visit d or when they always visit $d_{|\Delta|}$. This is the worst-case expected anonymity.

THEOREM 3. *A maximum of* $E[Y|X_D(u) = d]$ *occurs when either* $p^v_d = 1$ *for all* $v \neq u$ *or when* $p^v_{d_{|\Delta|}} = 1$ *for all* $v \neq u$.

PROOF SKETCH. Assume, following Lemma 2, that $(p^v)_{v \neq u}$ is such that $p^v_d = 1$ or $p^v_{d_{|\Delta|}} = 1$ for all $v \neq u$. Let $S \ni u$ be the set of users with unobserved inputs and $T \subseteq S$ be those that further have observed outputs. The expression for the expectation conditioned on S and T is maximized when $|\{s \in S - u : p^s_d = 1\}|$ is 0 or $|S|$. Which of the two is the maximum does not depend on S or T, and so a weighted sum over all such sets S and T is maximized when $|\{v \in U - u : p^v_d = 1\}|$ is 0 or $n-1$. $E[Y|X_D(u) = d \wedge X_O(u) = 0]$ is equal to such a sum. Equation 2 shows that, when $X_O(u) = 1$, the conditional expectation doesn't vary with the destination distributions $(p^v)_{v \neq u}$. □

3.2 Asymptotic Anonymity

The exact value of the maximum of $E[Y|X_D(u) = d]$ is not simple to express, but we can give a straightforward approximation for large user populations n. We focus on large n, because anonymity networks, and onion routing in particular, are understood to have the best chance at providing anonymity when they have many users. Furthermore, Tor is currently used by an estimated 200,000 people.

THEOREM 4. *When* $p^v_{d_{|\Delta|}} = 1$, *for all* $v \neq u$,

$$E[Y|X_D(u) = d] = b + b(1-b)p^u_d + (1-b)^2 p^u_d \left(\frac{1-b}{1-(1-p^u_{d_{|\Delta|}})b} + O\left(\sqrt{\frac{\log n}{n}}\right)\right) \quad (5)$$

PROOF. Let $f(n)$ represent the conditional expectation of the posterior probability in the case that u's input and output are not observed. Observe that in the case that u's output is observed $Y = 1$, because u is the only user that visits d. Thus $E[Y|X_D(u) = d] = b + b(1-b)p^u_d + (1-b)^2 f(n)$.

When u's input and output are unobserved, the value of Y depends on the number of other users with unobserved inputs, s, and the number of those s users with observed outputs, t. We can then express f as:

$$\begin{aligned}
f(n) &= E[Y|X_D(u)=d \wedge X_I(u)=0 \wedge X_O(u)=0] \\
&= \sum_{s=0}^{n-1}(1-b)^s b^{n-1-s}\binom{n-1}{s}\sum_{t=0}^{s} b^t(1-b)^{s-t}\binom{s}{t} \\
&\quad \cdot \frac{p_d^u\binom{s}{t}}{p^u_{d_{|\Delta|}}\binom{s}{t-1}+\binom{s}{t}} \\
&= \sum_{s=0}^{n-1}(1-b)^s b^{n-1-s}\binom{n-1}{s}\sum_{t=0}^{s} b^t(1-b)^{s-t}\binom{s}{t} \\
&\quad \cdot \frac{p_d^u(s-t+1)}{p^u_{d_{|\Delta|}}t+s-t+1}
\end{aligned}$$

Consider the inside sum. It is equal to the expected value of $g_0(s,T)=\frac{p_d^u(s-T+1)}{p^u_{d_{|\Delta|}}T+s-T+1}$, where $T \sim Bin(s,b)$. As s gets large, Chernoff bounds on the tails show that they contribute little to the expectation. Let $\varepsilon_0(s)$ be the terms of the sum for which t is more than $\sqrt{s\log s}$ from its expectation, $\mu_0 = bs$:

$$\begin{aligned}
\varepsilon_0(s) &= \sum_{t:|t-\mu_0|>\sqrt{s\log s}} b^t(1-b)^{s-t}\binom{s}{t}g_0(s,t) \\
&\leq \sum_{t:|t-\mu_0|>\sqrt{s\log s}} b^t(1-b)^{s-t}\binom{s}{t} \qquad \text{as } g_0 \leq 1 \\
&\leq 2e^{-c_0\log s/b} \qquad \text{for } c_0 < 1/2 \text{ and large } s \\
&\qquad\qquad\qquad\quad \text{by Chernoff's inequality} \\
&= s^{-c_1} \qquad \text{for some } c_1 > 1/2
\end{aligned}$$

Now we look at how much the values of g_0 inside the tail differ from the value of g_0 at its expectation. Let $\varepsilon_1(s,t) = g_0(s,t) - g_0(s,\mu_0)$ be this difference. It can be shown through direct calculation that $\frac{\partial\varepsilon_1}{\partial t} \leq 0$ and $\frac{\partial^2\varepsilon_1}{\partial t^2} \leq 0$. Therefore for values of t that are within $\sqrt{s\log s}$ of μ_0:

$$\begin{aligned}
|\varepsilon_1(s,t)| &\leq \left|\varepsilon_1\left(s,\mu_0+\sqrt{s\log s}\right)\right| \\
&= O\left(\sqrt{\frac{\log s}{s}}\right) \qquad \text{by direct calculation}
\end{aligned}$$

Looking now at the outside sum we have:

$$\begin{aligned}
f(n) = &\sum_{s=0}^{n-1}(1-b)^s b^{n-1-s}\binom{n-1}{s} \\
&\left[\frac{p_d^u(s(1-b)+1)}{s\left(1-\left(1-p^u_{d_{|\Delta|}}\right)b\right)+1}\right. \\
&\left. +O(s^{-c_1})+O\left(\sqrt{\frac{\log s}{s}}\right)\right]
\end{aligned}$$

This is the expectation of a function of a binomially distributed random variable with mean $\mu_1 = (1-b)(n-1)$. Let $\varepsilon_2(n)$ be the parts of this sum that are greater than $k(n-1)$ from μ_1, $k < \min(b, 1-b)$:

$$\begin{aligned}
\varepsilon_2(n) &\leq \sum_{s:|s-\mu_1|>k(n-1)}(1-b)^s b^{n-1-s}\binom{n-1}{s} \\
&\leq 2e^{-c_2k^2(n-1)/(1-b)} \qquad \text{for some } c_2 > 0 \\
&\qquad\qquad\qquad\qquad\quad \text{by Chernoff's inequality} \\
&= O(e^{-c_3n}) \qquad \text{where } c_3 = c_2k^2/(1-b)
\end{aligned}$$

Let g_1 be the non-vanishing inner term of $f(n)$:

$$g_1(s) = \frac{p_d^u(s(1-b)+1)}{s\left(1-\left(1-p^u_{d_{|\Delta|}}\right)b\right)+1}$$

As s grows it approaches a limit of:

$$g_1^* = \frac{p_d^u(1-b)}{1-\left(1-p^u_{d_{|\Delta|}}\right)b}$$

Let $\varepsilon_3(s) = g_1(s) - g_1^*$ be the difference from this limit. Direct calculation shows that $\frac{d\varepsilon_3}{ds} \leq 0$ and $\frac{d^2\varepsilon_3}{ds^2} \geq 0$. Therefore for values of s within $k(n-1)$ of μ_1:

$$\begin{aligned}
|\varepsilon_3(s)| &\leq \varepsilon_3(\mu_1 - k(n-1)) \\
&= O(1/n)
\end{aligned}$$

Now we can show the desired asymptotic expression for the entire sum:

$$\begin{aligned}
f(n) &= O(e^{-c_3n}) + \sum_{s=\mu_1-k(n-1)}^{\mu_1+k(n-1)}(1-b)^s b^{n-1-s}\binom{n-1}{s} \\
&\quad \cdot\left[g_1^* + \varepsilon_3(s) + O(s^{-c_1}) + O\left(\sqrt{\frac{\log s}{s}}\right)\right] \\
&= g_1^* + O(e^{-c_3n}) + O(1/n) + O(n^{-c_1}) \\
&\quad + O\left(\sqrt{\frac{\log n}{n}}\right) \\
&= \frac{p_d^u(1-b)}{1-\left(1-p^u_{d_{|\Delta|}}\right)b} + O\left(\sqrt{\frac{\log n}{n}}\right)
\end{aligned}$$

□

Theorem 5. *When $p_d^v = 1$, for all $v \neq u$,*

$$\begin{aligned}
&E[Y|X_D(u)=d] = b^2 + b(1-b)p_d^u + \\
&\qquad (1-b)\frac{p_d^u}{1-(1-p_d^u)b} + O\left(\sqrt{\frac{\log n}{n}}\right) \qquad (6)
\end{aligned}$$

Proof. The proof of this theorem is similar to that of Theorem 4.

Let $f(n)$ represent the conditional expectation of the posterior probability in the case that u's input is not observed. Thus $E[Y|X_D(u)=d] = b^2 + b(1-b)p_d^u + (1-b)f(n)$.

We can express f as:

$$\begin{aligned}
f(n) &= E[Y|X_D(u) = d \wedge X_I(u) = 0] \\
&= \sum_{s=0}^{n-1}(1-b)^s b^{n-1-s}\binom{n-1}{s} \\
&\quad \cdot \sum_{t=0}^{s+1} b^t(1-b)^{s+1-t}\binom{s+1}{t}\frac{p_d^u\binom{s}{t} + p_d^u\binom{s}{t-1}}{\binom{s}{t} + p_d^u\binom{s}{t-1}} \\
&= \sum_{s=0}^{n-1}(1-b)^s b^{n-1-s}\binom{n-1}{s} \\
&\quad \cdot \sum_{t=0}^{s+1} b^t(1-b)^{s+1-t}\binom{s+1}{t}\frac{p_d^u(s+1)}{s-(1-p_d^u)t+1}
\end{aligned}$$

As s gets large, Chernoff bounds on the tails of the inner sum show that they contribute little to the expectation. Let $g_0(s,t) = \frac{p_d^u(s+1)}{s-(1-p_d^u)t+1}$. Let $\varepsilon_0(s)$ be the terms of the inner sum for which t is more than $\sqrt{s \log s}$ from $\mu_0 = b(s+1)$:

$$\begin{aligned}
\varepsilon_0(s) &= \sum_{t:|t-\mu_0|>\sqrt{s\log s}} b^t(1-b)^{s+1-t}\binom{s+1}{t} g_0(s,t) \\
&\leq s^{-c_1} \qquad \text{for some } c_1 > 1/2 \\
&\qquad\qquad \text{by Chernoff's inequality}
\end{aligned}$$

Let $\varepsilon_1(s,t) = g_0(s,t) - g_0(s,\mu_0)$. It can be shown through direct calculation that $\frac{\partial\varepsilon_1}{\partial t} \geq 0$ and $\frac{\partial^2\varepsilon_1}{\partial t^2} \geq 0$. Therefore for values of t that are within $\sqrt{s\log s}$ of μ_0:

$$\begin{aligned}
|\varepsilon_1(s,t)| &\leq \varepsilon_1\left(s, \mu_0 + \sqrt{s\log s}\right) \\
&= O\left(\sqrt{\frac{\log s}{s}}\right)
\end{aligned}$$

Looking now at the outside sum we have:

$$\begin{aligned}
f(n) &= \sum_{s=0}^{n-1}(1-b)^s b^{n-1-s}\binom{n-1}{s} \\
&\quad \cdot \left[\frac{p_d^u}{1-(1-p_d^u)b} + O(s^{-c_1}) + O\left(\sqrt{\frac{\log s}{s}}\right)\right]
\end{aligned}$$

Let $\varepsilon_2(n)$ be the parts of this sum that are greater than $k(n-1)$ from $\mu_1 = (n-1)(1-b)$, $k < \min(b, 1-b)$:

$$\begin{aligned}
\varepsilon_2(n) &\leq \sum_{s:|s-\mu_1|>k(n-1)} (1-b)^s b^{n-1-s}\binom{n-1}{s} \\
&\leq 2e^{-c_2k^2(n-1)/(1-b)} \qquad \text{for some } c_2 > 0 \\
&\qquad\qquad \text{by Chernoff's inequality} \\
&= O(e^{-c_3 n}) \qquad \text{where } c_3 = c_2k^2/(1-b)
\end{aligned}$$

Now we can show the desired asymptotic expression for the entire sum:

$$\begin{aligned}
f(n) &= O(e^{-c_3n}) \\
&\quad + \sum_{s=(1-b-k)(n-1)}^{(1-b+k)(n-1)} (1-b)^s b^{n-1-s}\binom{n-1}{s} \\
&\quad \cdot \left[\frac{p_d^u}{1-(1-p_d^u)b} + O(s^{-c_1}) + O\left(\sqrt{\frac{\log s}{s}}\right)\right] \\
&= \frac{p_d^u}{1-(1-p_d^u)b} + O\left(\sqrt{\frac{\log n}{n}}\right)
\end{aligned}$$

□

To determine which distribution is the worst case for large n, simply examine the difference between the limits of the expressions in Theorems 4 and 5. It is clear from this that the worst-case distribution is $p_d^v = 1, \forall v \neq u$, only when $p^u_{d_{|\Delta|}} \geq \frac{(1-b)(1-p_d^u)^2}{p_d^u(1+b)-b}$. This happens when $p_d^u \geq 1/2$ and $p^u_{d_{|\Delta|}}$ is near $1 - p_d^u$. For $p^u_{d_{|\Delta|}}$ small, which we would expect as it must be less than $1/|\Delta|$, the worst-case distribution is $p^v_{d_{|\Delta|}} = 1, \forall v \neq u$. In this case the expected assigned probability is about $b + (1-b)p_d^u$. This can be viewed as decreasing the "innocence" of u from $1-p_d^u$ to $(1-b)(1-p_d^u)$. It is also equal to the lower bound on anonymity in onion routing when the adversary controls a fraction $\sqrt{b}$ of the network.

4. TYPICAL DISTRIBUTIONS

It is unlikely that users of onion routing will ever find themselves in the worst-case situation. The necessary distributions just do not resemble what we expect user behavior to be like in any realistic use of onion routing. Our worst-case analysis may therefore be overly pessimistic. To get some insight into the anonymity that a typical user of onion routing can expect, we consider a more realistic set of users' destination distributions in which each user selects a destination from a common Zipfian distribution. This model of user behavior is used by Shmatikov and Wang [25] to analyze relationship anonymity in mix networks and is motivated by observations that the popularity of sites on the web follows a Zipfian distribution. Our results show that a user's expected assigned probability is close to $b^2 + (1-b^2)p_d^u$ for large populations, which is the best that can be expected when there is a b^2 probability of total compromise.

Let each user select his destination from a common Zipfian distribution p: $p_{d_i} = 1/(\mu i^s)$, where $s > 0$ and $\mu = \sum_{i=1}^{|\Delta|} 1/i^s$. It turns out that the exact form of the distribution doesn't matter as much as the fact that it is common among users.

Theorem 6. *When $p^v = p^w$, for all $v, w \in U$,*

$$E[Y|X_D(u) = d] = b^2 + (1-b^2)p_d^u + O(1/n)$$

Proof. Let p be the common destination distribution.

The expected assigned probability can be expressed as:

$$\begin{aligned}
&E[Y|X_D(u)=d] = b^2 + b(1-b)p_d^u + \\
&\quad (1-b)\sum_{s=1}^{n} b^{n-s}(1-b)^{s-1}\binom{n-1}{s-1}\sum_{t=0}^{s}(1-b)^{s-t}b^t \\
&\quad \cdot \left[\binom{s-1}{t-1}\sum_{\Delta\in D^t:\Delta_1=d}\prod_{i=2}^{t} p_{\Delta_i}\psi(s,\Delta) + \right. \\
&\quad \left. \binom{s-1}{t}\sum_{\Delta\in D^t}\prod_{i=1}^{t} p_{\Delta_i}\psi(s,\Delta)\right]
\end{aligned}$$

Here, s represents the size of the set of users with unobserved inputs, t represents the size of the subset of those s users that also have observed outputs, Δ represents the t observed destinations, and $\psi(s,\Delta)$ is the posterior probability.

Let $\Delta_d = |\{x \in \Delta : x = d\}|$. The posterior probability ψ can be expressed simply as:

$$\begin{aligned}
\psi(s,\Delta) &= \frac{\Delta_d(s-1)^{|\Delta|-1} + p_d(s-1)^{|\Delta|}}{s^{|\Delta|}} \\
&= (\Delta_d + p_d(s-t))/s
\end{aligned}$$

The sum $\sum_{\Delta\in D^t:\Delta_1=d}\prod_{i=2}^{t} p_{\Delta_i}\psi(s,\Delta)$ calculates the expectation for ψ conditioned on s and t. The expression for ψ shows that this depends linearly on the expected value of Δ_d. This expectation is simply $1 + p_d(t-1)$, because one destination in this case is always t, and each of the other $t-1$ is d with probability p_d. The sum $\sum_{\Delta\in D^t}\prod_{i=1}^{t} p_{\Delta_i}\psi(s,\Delta)$ similarly depends linearly on the expectation of Δ_d, which in this case is $p_d t$.

With this observation, it is a straightforward calculation to show that the inner sum over t is simply:

$$b\frac{p_d(s-1)+1}{s} + (1-b)p_d$$

We insert this into the larger sum and simplify:

$$\begin{aligned}
E[Y|X_D(u)=d] &= b^2 + b(1-b)p_d^u + (1-b) \\
&\quad \cdot \sum_{s=1}^{n} b^{n-s}(1-b)^{s-1}\binom{n-1}{s-1} \\
&\quad \cdot \left[b\frac{p_d(s-1)+1}{s} + (1-b)p_d\right] \\
&= b^2 + (1-b^2)p_d^u + O(1/n)
\end{aligned}$$

□

5. CONCLUSIONS AND FUTURE WORK

We expect each user of an anonymity network to have a pattern of use. In order to make guarantees to the user about his anonymity, we need to take this into account when modeling and analyzing the system, especially in light of previous research that indicates that an adversary can learn these usage patterns given enough time.

We perform such an analysis on onion routing. Onion routing is a successful design used, in the form of the Tor system, by hundreds of thousands of people to protect their privacy, but, because it was designed to be practical and because theory in this area is still relatively young, the formal analysis of its privacy properties has been limited. Our analysis is ultimately based on a formal protocol specification, but we show that it can be captured with a simple black-box model that should lend itself to the analysis of other anonymity protocols. We investigate the relationship anonymity of users and their destinations in this model and measure it with the probability that the adversary assigns to the correct destination of a given user after observing the network.

We first consider the worst-case set of user behaviors to give an upper bound on anonymity. We show that a user's anonymity is worst either when all other users choose destinations he is unlikely to visit, because that user becomes unique and identifiable, or when that user chooses a destination that all other users prefer, because the adversary mistakes the group's choices for the user's choice. This worst-case anonymity with an adversary that controls a fraction b of the routers is comparable to the best-case anonymity against an adversary that controls a fraction $\sqrt{b}$.

The worst case is unlikely to be the case for any users; so we investigate anonymity under a more reasonable model of user behavior suggested in the literature. In it, users select destinations from a common Zipfian distribution. Our results show that, in this case and in any case with a common distribution, the expected anonymity tends to the best possible, *i.e.* the adversary doesn't usually gain that much knowledge from the other users' actions.

Future work includes extending this analysis to other types of anonymity (such as sender anonymity), extending it to other anonymity networks, and learning more about the belief distribution of the adversary than just its mean. A big piece of the attack we describe is in learning the users' destination distribution, about which only a small amount of research, usually on simple models, has been done. The speed with which an adversary can perform this stage of the attack is crucial in determining the validity of our attack model and results.

In response to analyses such as that of Øverlier and Syverson [18], the current Tor design includes entry guards by default for all circuits. Roughly, this means that, since about January 2006, each Tor client selects its first onion router from a small set of nodes that it randomly selects at initialization. The rationale is that communication patterns of individuals are what need to be protected. If an entry guard is compromised, then the percentage of compromised circuits from that user is much higher. But, without entry guards, it appears that whom that user communicates with and even at what rate can be fairly quickly learned by an adversary owning a modest percentage of the Tor nodes anyway. If no entry guard is compromised, then no circuits from that user will ever be linked to him. However, if a user expects to be targeted by a network adversary that can control nodes, he can expect his entry guards ultimately to be attacked and possibly compromised. If the destinations he chooses that are most sensitive are rarely contacted, he may thus be better off choosing first nodes at random. How can we know which is better? Extending our analysis to include entry guards will allow us to answer or at least further illuminate this question.

Our model also assumes that client connections to the network are such that the initial onion router in a circuit can tell that it is initial for that circuit. This is true for the overwhelming majority of traffic on the Tor network today, because most users run clients that are not also onion routers.

However, for circuits that are initiated at a node that runs an onion router, a first node cannot easily tell whether it is the first node or the second—without resorting to other attacks of unknown efficacy, e.g., monitoring latency of traffic moving in each direction in response to traffic moving in the other direction. Thus, that initiating edge of the black box is essentially fuzzy. Indeed, this was originally the only intended configuration of onion routing for this reason [9]. The addition of clients that do not also function as routers was a later innovation that was added to increase usability and flexibility [20, 26]. Similarly, peer-to-peer designs such as Crowds [21] and Tarzan [8] derive their security even more strongly from the inability of the first node to know whether it is first or not. Thus, extending our model and analysis to this case will make it still more broadly applicable.

6. REFERENCES

[1] P. Boucher, A. Shostack, and I. Goldberg. Freedom systems 2.0 architecture. White paper, Zero Knowledge Systems, Inc., 2000.

[2] J. Camenisch and A. Lysyanskaya. A formal treatment of onion routing. In *Proceedings of CRYPTO 2005*, pages 169–187, 2005.

[3] G. Danezis. Statistical disclosure attacks: Traffic confirmation in open environments. In *Proceedings of Security and Privacy in the Age of Uncertainty (SEC 2003)*, pages 421–426, 2003.

[4] G. Danezis and A. Serjantov. Statistical disclosure or intersection attacks on anonymity systems. In *Proceedings of 6th Information Hiding Workshop (IH 2004)*, pages 293–308, 2004.

[5] C. Díaz, S. Seys, J. Claessens, and B. Preneel. Towards measuring anonymity. In *Proceedings of the 2nd Privacy Enhancing Technologies Workshop (PET 2002)*, pages 54–68, 2002.

[6] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th USENIX Security Symposium*, pages 303–320, 2004.

[7] J. Feigenbaum, A. Johnson, and P. Syverson. A model of onion routing with provable anonymity. In *Proceedings of the 11th Financial Cryptography and Data Security Conference (FC 2007)*, 2007.

[8] M. J. Freedman and R. Morris. Tarzan: A peer-to-peer anonymizing network layer. In *Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS 2002)*, pages 193–206, 2002.

[9] D. M. Goldschlag, M. G. Reed, and P. F. Syverson. Hiding Routing Information. In *Proceedings of the 1st Information Hiding Workshop (IH 1996)*, pages 137–150, 1996.

[10] J. Y. Halpern and K. R. O'Neill. Anonymity and information hiding in multiagent systems. *Journal of Computer Security*, 13(3):483–514, 2005.

[11] D. Hughes and V. Shmatikov. Information hiding, anonymity and privacy: A modular approach. *Journal of Computer Security*, 12(1):3–36, 2004.

[12] D. Kesdogan, D. Agrawal, and S. Penz. Limits of anonymity in open environments. In *Proceedings of the 5th Information Hiding Workshop (IH 2002)*, pages 53–69, 2002.

[13] D. Kesdogan, J. Egner, and R. Büschkes. Stop-and-go MIXes: Providing probabilistic anonymity in an open system. In *Proceedings of the 2nd Information Hiding Workshop (IH 1998)*, pages 83–98, 1998.

[14] P. Lincoln, P. Porras, and V. Shmatikov. Privacy-preserving sharing and correlation of security alerts. In *Proceedings of the 13th USENIX Security Symposium*, pages 239–254, 2004.

[15] N. Mathewson and R. Dingledine. Practical traffic analysis: Extending and resisting statistical disclosure. In *Proceedings of the 4th Privacy Enhancing Technologies workshop (PET 2004)*, pages 17–34, 2004.

[16] S. Mauw, J. Verschuren, and E. de Vink. A formalization of anonymity and onion routing. In *Proceedings of the 9th European Symposium on Research in Computer Security (ESORICS 2004)*, pages 109–124, 2004.

[17] Number of running tor routers. http://www.noreply.org/tor-running-routers/, May 2007.

[18] L. Øverlier and P. Syverson. Locating hidden servers. In *Proceedings of the 17th IEEE Symposium on Security and Privacy (S&P 2006)*, pages 100–114, 2006.

[19] A. Pfitzmann and M. Hansen. Anonymity, unobservability, and pseudonymity: A consolidated proposal for terminology. Draft, July 2000.

[20] M. Reed, P. Syverson, and D. Goldschlag. Anonymous connections and onion routing. *IEEE Journal on Selected Areas in Communications*, 16(4):482–494, 1998.

[21] M. Reiter and A. Rubin. Crowds: Anonymity for web transactions. *ACM Transactions on Information and System Security*, 1(1):66–92, 1998.

[22] S. Schneider and A. Sidiropoulos. CSP and anonymity. In *Proceedings of the 1st European Symposium on Research in Computer Security (ESORICS 1996)*, pages 198–218, 1996.

[23] A. Serjantov and G. Danezis. Towards an information theoretic metric for anonymity. In *Proceedings of the 2nd Privacy Enhancing Technologies Workshop (PET 2002)*, pages 41–53, 2002.

[24] V. Shmatikov. Probabilistic model checking of an anonymity system. *Journal of Computer Security*, 12(3-4):355–377, 2004.

[25] V. Shmatikov and M.-H. Wang. Measuring relationship anonymity in mix networks. In *Proceedings of the 5th ACM Workshop on Privacy in the Electronic Society (WPES 2006)*, pages 59–62, 2006.

[26] P. Syverson, M. Reed, and D. Goldschlag. Onion routing access configurations. In *Proceedings of the DARPA Information Survivability Conference and Exposition (DISCEX 2000)*, pages 34–40, 2000.

[27] P. Syverson, G. Tsudik, M. Reed, and C. Landwehr. Towards an Analysis of Onion Routing Security. In *Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, pages 96–114, 2000.

[28] P. F. Syverson and S. G. Stubblebine. Group principals and the formalization of anonymity. In *Proceedings of the 1st World Congress on Formal Methods (FM'99), Vol. I*, pages 814–833, 1999.

[29] G. Tóth, Z. Hornák, and F. Vajda. Measuring anonymity revisited. In *Proceedings of the 9th Nordic Workshop on Secure IT Systems*, pages 85–90, 2004.

[30] M. K. Wright, M. Adler, B. N. Levine, and C. Shields. The predecessor attack: An analysis of a threat to anonymous communications systems. *ACM Transactions on Information and Systems Security*, 7(4):489–522, 2004.